

Petra Storjohann

COMPUTERGESTÜTZTE LESARTENDISAMBIGUIERUNG

Abstract

Mit diesem Papier sollen Lexikografinnen an ein Automatisierungstool der Textanalyse innerhalb der Korpuslinguistik herangeführt werden. Das am IDS entwickelte statistische Recherche- und Analysewerkzeug *Cosmas* bietet neue Zugänge zur Gewinnung semantischer Informationen über Wörter. Die Nutzungsmöglichkeiten dieses Instrumentariums für die Lesartendisambiguierung von Lexemen und deren Verifizierung mittels Kollokations- und Kontextanalyse werden erläutert, und anhand des Beispiels *cool* wird gezeigt, inwieweit sich semantische Informationen durch automatische Statistik extrahieren lassen. Dabei wird auf die Vor- und Nachteile der computerbasierten Analyse eingegangen. Darüber hinaus wird dargestellt, wie empirische lexikografische Disambiguierung modellgeleitet validiert werden kann. Um die Unterschiede zwischen herkömmlichen Beschreibungsmöglichkeiten und neuen statistischen Verfahren zu verdeutlichen, werden die Lesarten zu *cool*, wie sie im Duden GWDS (2000) dargestellt sind, mit den identifizierten Lesarten der Analyse mit *Cosmas* verglichen.

The purpose of this paper is to introduce lexicographers to a computational tool for automatic content analysis in corpus linguistics. *Cosmas*, an efficient statistical search and text analysis tool developed at the IDS offers new ways of obtaining semantic information about words. I shall demonstrate the enhanced disambiguation techniques provided by this instrument for analysing polysemous lexemes and show how the results can be validated with the help of collocational and contextual analyses. This will be exemplified by the word *cool*, which illustrates how far computational and statistical methods are able to generate semantic information. Furthermore, I shall seek to emphasise the advantages and disadvantages of statistical computer-based identification. I shall also discuss how empirical lexicographic disambiguation can be validated within a theoretical framework. In order to illuminate the differences between traditional semantic description and the use of computational tools for automatic analysis, I shall draw a comparison with the dictionary entry in the Duden GWDS (2000) for the lexeme *cool*.

1. Korpusbasierte lexikografische Lesartenidentifikation

Elektronische Textkorpora als Arbeitsgrundlage haben seit einiger Zeit die Arbeit der Lexikografinnen¹ verändert. Die Vorteile einer digitalisierten Textsammlung sah man in erster Linie in der Möglichkeit, schnell treffende Belege zu finden, um damit die durch Introspektion gewonnenen Beschreibungen zu stützen.² Nach wie vor erweist es sich als problematisch, Korpora stärker in den Prozess der Erkenntnisgewinnung einzubeziehen. Oft wird der Vorteil des Einsatzes von Korpora für die Entwicklung neuer Ansätze, die dazu dienen, zu neuen Beschreibungsmöglichkeiten zu gelangen, nicht bedacht. Auf die Möglichkeit, neue Wege hin zu objektiveren Resultaten zu finden, macht Sinclair (1991, S. 36) zu Recht aufmerksam:

It is the possibility of new approaches, new kinds of evidence, and new kinds of description. Here, the objectivity and surface validity of computer techniques become an asset rather than a liability. Without relinquishing our intuition, of course, we try to find explanations that fit the evidence, rather than adjusting the evidence to fit the pre-set explanation.

¹ Vgl. Scholze-Stubenrecht (2002, S. 233) und Scholze-Stubenrecht (2001, S. 43) zu Veränderungen der lexikografischen Arbeit in der Dudenredaktion. Vgl. auch Sinclair (1991, S. 1-8).

² Vgl. Haß-Zumkehr (2002, S. 45).

Obwohl die Korpustechnologie im Bereich Semantik ihre Vorteile bereits geltend machen konnte, gibt es eigentlich kaum automatische Werkzeuge, die Korpusdaten semantisch analysieren und auswerten können (vgl. Scholze-Stubenrecht 2002, S. 234).³ Wie Jones (2002) hervorhebt, stehen viele WissenschaftlerInnen automatisierten Verfahren zur Erkennung semantischer Strukturen skeptisch gegenüber. Dabei können Recherche- und Analysetools, die mit statistischen Methoden arbeiten, helfen, semantisch Auffälliges innerhalb großer Textmengen aufzuspüren und regelhafte Strukturen zu erkennen, indem sie die Rechercheergebnisse systematisiert und strukturiert den LexikografInnen zur weiteren Bearbeitung zur Verfügung stellen. Solche Rechercheergebnisse können unter anderem bei der semantischen Beschreibung, wie sie in Wörterbüchern vorgenommen wird, stärker genutzt werden, da sie eine Vielfalt an Informationen für eine empirisch basierte Beschreibung liefern. Die Erkenntnis, dass Korpora der Lexikografie die Möglichkeit bieten, empirisch fundierte Erkenntnisse zu gewinnen, findet in der Praxis erst in jüngerer Zeit Resonanz.⁴

Das IDS verfügt über den größten deutschsprachigen Datenbestand in Form verschiedener Korpora.⁵ Diese sind mittels „COSMAS-Korpus-Recherchesystem“ (künftig als *Cosmas*⁶ zitiert) recherchierbar, einem Recherche- und Analysewerkzeug, welches den linguistischen Zugang zur digitalen Textsammlung des IDS herstellt. *Cosmas* wurde in den 90-er Jahren entwickelt und wird im Projekt *Wissen über Wörter / elexiko*⁷ (künftig als *elexiko* bezeichnet) unter anderem für die lexikografische Lesartendisambiguierung genutzt. Im Folgenden wird das theoriegeleitete und computerbasierte Verfahren der Lesartenidentifikation mittels *Cosmas* beschrieben und der Gewinn gegenüber traditionellen lexikografischen Verfahren, wie sie u.a. von der Dudenredaktion benutzt werden, verdeutlicht. Für die Lesartendisambiguierung ist eine Komponente von *Cosmas*, die

³ Es existieren zwei Arten von automatischen Werkzeugen. Zum einen gibt es Tools, die auf mathematisch-statistischen Grundlagen arbeiten, ohne sprachwissenschaftlichen Zwecken zu dienen, zum anderen gibt es Werkzeuge, die für die linguistische Inhaltserschließung textlicher Massendaten entwickelt wurden. Besonders innerhalb der englischen Linguistik wurden bereits elaborierte Systeme für das semantische Information Retrieval entwickelt, die in Zusammenhang mit Korpora wie dem British National Corpus oder dem Brown Corpus arbeiten. Diese sind u.a. BNCWeb, SARA/XARA, LDC's collocation calculator, WordCruncher und WordSmith.

⁴ Vgl. Dänisches Wörterbuch DDO – Den Danske Ordbog <http://www.dsl.dk/ddo.html>) und Benedict – The New Intelligent Dictionary (<http://www.comp.lancs.ac.uk/computing/research/ucrel/projects.html#benedict>).

⁵ Informationen zur Korpuszusammensetzung unter: <http://www/kt/projekte/korpora/archiv.html> Vgl. auch Haß-Zumkehr (2003).

⁶ Mit der Kurzform *Cosmas* (Corpus Search, Management and Analysis System) wird auf den vollständigen Namen COSMAS-Korpus-Recherchesystem Bezug genommen. Das Recherche- und Analysetool steht zur Nutzung der Öffentlichkeit im Internet zur Verfügung (<http://cosmas2/>).

⁷ Näheres unter: <http://www/elexiko/>.

Software „Statistische Kollokationsanalyse und Clustering“⁸ (im Folgenden als Kollokationsanalyse oder Kookkurrenzanalyse zitiert)⁹ von besonderem Interesse.

Die Analyse der Kookkurrenzen stellt einen neuen Zugang zur Gewinnung semantischer Information dar, auf deren Grundlage LexikografInnen zur Zerlegung semantischer Strukturen, zur Erkennung von Verwendungsmöglichkeiten eines Wortes und damit zur Lesartendisambiguierung gelangen. Die statistische Auswertung der Kookkurrenzen erlaubt die Identifikation von Lesarten und semantische Beschreibungstiefen, die zum Teil erheblich von denen des entsprechenden Dudeneintrages des Dudens GWDS (2000)¹⁰ abweichen. *Elexiko* hat ein Disambiguierungsverfahren entwickelt, welches auf empirischen und theoretischen Grundlagen basiert. Das praktische Verfahren der Lesartenidentifikation wird modellgeleitet durchgeführt und im Folgenden anhand des Beispielllexems *cool* in Abschnitt 2. und 3. erläutert.

2. Praktisches Verfahren der Lesartenidentifikation

Die korpusbasierte Disambiguierung einzelner Lesarten und die dafür notwendige vollständige semantische Analyse eines Lexems beginnt mit der Kollokationsanalyse. Ziel der Kollokationsanalyse ist es, wichtige lexikalische Strukturen und Muster, die für ein Lexem bedeutungskonstitutiv sind, durch die Ermittlung von signifikanten¹¹ Kookkurrenzen zu erhalten. Die Kollokationsanalyse ermittelt statistisch das Miteinandervorkommen von Lexemen in Kontexten¹² und bietet ein Ordnungssystem an, indem sie die Kontexte nach unterschiedlichen Aspekten systematisiert und klassifiziert. Den LexikografInnen wird damit eine semantisch vorstrukturierte Umgebung des gesuchten Wortes zur Verfügung gestellt. Wie auch Sinclair (1991) hervorhebt, liegt der große Nutzen der Kollokationsanalyse in der Organisation und Strukturierung der Kontexte eines Lexems.

Ein kurzer Ausschnitt aus der Kollokationsliste des Lexems *cool* soll die vom Programm vorgenommene Anordnung¹³ illustrieren:

⁸ Die Software „Statistische Kollokationsanalyse und Clustering“, auf die in Zusammenhang mit *Cosmas* zugegriffen werden kann, wurde auf der Basis statistischer Methoden von Cyril Belica (1995-2002) am IDS entwickelt und kann seit 1995 kostenlos online genutzt werden. (siehe auch Informationen zu Urheberrechten unter <http://www.fkt-projekte/methoden/ka.html>). Eine genaue Erläuterung der Verfahrensgrundlagen soll an dieser Stelle nicht erfolgen. Ziel dieses Artikels soll es sein, das Instrumentarium in seiner Anwendung zu beschreiben.

⁹ Der Begriff *Kollokation* geht hier über die Vorstellung der strengen syntaktischen binären Relation hinaus und lehnt sich an den Begriff *collocation* der Korpuslinguistik an. Kookkurrenz wird hier als Miteinandervorkommen von Wörtern innerhalb eines Kontextes verstanden. Vgl. auch Steyer (2000).

¹⁰ Duden GWDS (2000): *Das große Wörterbuch der deutschen Sprache*. 10 Bände auf CD-Rom.

¹¹ Unter Signifikanz wird statistische Signifikanz verstanden. Sie wird durch den Log-Likelihood-Ratio-Wert, welcher den Grad der lexikalischen Kohäsion bestimmt, zum Ausdruck gebracht.

¹² Vgl. Lawson (2000).

¹³ Die Anordnung erfolgt nach Grad der lexikalischen Kohäsion.

BelegNr	Gamma	Kollokatoren	Häufigkeit
2676+46:	369	Sonnenbrille	46
2722+46:	363	geil	46
2768+52:	355	total	52
2820+47:	352	einfach sein	47
2867+61:		einfach	61
2928+63:	345	Jung	63
2991+231:	334	finden	231
3222+1:	310	stilvoll	1
3223+20:	308	Säue	20
3243+2:	306	Mister Planke 16.45	2
3245+5:		Mister Planke	5
3250+6:		Mister 16.45	6
3256+20:		Mister	20
3276+23:	302	Water	23
3299+1491:	300	sein	1491
4790+10:	299	voll Angus	10
4800+35:		voll	35
4835+1:	297	Session	1
4836+2:	294	basieren	2
4838+3:	292	Baby	3
4841+29:	274	clever	29

Eine vollständige Liste aller Kollokatoren erschreckt das ungeübte Auge, kann sie doch bei hochfrequenten Lexemen mehrere Seiten lang sein. Aber bei gründlicher Betrachtung werden die Beziehungen zwischen dem zu untersuchenden Lexem und seinen lexikalischen Mitspielern sichtbar. Die von der Kollokationsanalyse vorgenommene Kookkurrenzauswertung erstellt in Abhängigkeit von Wortart und Einstellungen der Parameter eine Liste der benachbarten okkasionellen und usuellen lexikalischen Mitspieler.¹⁴ Die Ermittlung von Frequenz, statistischer Signifikanz und Abstand der Mitspieler zum gesuchten Lexem verdeutlicht die lexikalisch-semantische Kohäsion, die für Lexikografinnen für die Lesartendisambiguierung von besonderem Interesse ist. Die extrahierten Kollokatoren liefern aufgrund ihrer Vielfalt ein ganzes Netz an Informationen. Sie zeigen unterschiedlichste sprachliche Zusammenhänge wie paradigmatische Wortschatzvernetzungen, idiomatische Wendungen, thematische Bezüge zu Sachgebieten und zur Diskursgeschichte sowie syntaktische Konstruktionen. Neben usuellen Wortverbindungen werden auch umfangreichere syntagmatische Kombinationen (Phraseologismen, Redewendungen, Sprichwörter, kommunikative Formeln, Funktionsverbgefüge) erfasst, deren Analyse ebenfalls einen Beitrag zu den Verwendungsweisen eines Lexems leistet.

Die semantische Untersuchung der von der Kollokationsanalyse vorstrukturierten Kookkurrenzen bietet unterschiedliche Perspektiven und Beschreibungsmöglichkeiten in Hinblick auf die Verwendung des zu untersuchenden Wortes. Variante und invariante Strukturen werden aufgespürt sowie unterschiedliche Verwendungsmuster verdeutlicht. Mit ihnen können Kontexte aufgezeigt und differenzierte semantische und pragmatische Angaben zu eben diesen Kontexten gegeben werden. Wie Sinclair (1991, S. 61) betont,

¹⁴ Zu okkasionellen und usuellen Verbindungen vgl. Steyer (2000).

besteht der Nutzen der Kollokationsanalyse in der statistischen Beweisführung über die Existenz von gruppierbaren linguistischen Strukturen und Bedeutungen.

The evidence is initially statistical, but is subjected to the application of a linguistic principle such as the alignment of sense and structure. The result should be a powerful evaluative argument.

Die Kollokationsanalyse deckt lexikalische Strukturen durch Gruppierungen von signifikanten Mitspielern auf. In Bezug auf die Disambiguierung eines Lexems können die Ergebnisse aber nur als Indikatoren gewertet werden, deren Verifizierung nur durch eine von Lexikografinnen selbst vorgenommene Überprüfung des Kontextes¹⁵ des Kollokators durchgeführt werden kann. Diese Notwendigkeit ergibt sich aus mehreren Gründen: Zum einen kann die Kollokationsanalyse nicht zwischen homonymen Lexemen unterscheiden und verweist daher auf sämtliche Mitspieler, die einer lexikalischen Form zugeordnet sind; zum anderen entstehen fehlgeleitete Kollokationszuweisungen durch Fehler der Lemmatisierungssoftware. Um auf irrelevante Kookkurrenzen aufmerksam zu werden, bedarf es einer näheren Betrachtung des kontextuellen Umfeldes jedes einzelnen Mitspielers. Eine klare Abgrenzung der aussagekräftigen Kollokatoren von Partnern, die für die Disambiguierung irrelevant sind, kann nur durch die Analyse der Referenzbereiche sowie durch die Analyse paradigmatischer und syntagmatischer Muster aus den Kollokatorenkontexten gewährleistet werden. Entsprechende Kontextausschnitte werden durch Anklicken des entsprechenden Kollokators angezeigt, wodurch die nächste Analyseebene, die KWIC-Ebene¹⁶ (Key Word In Context) geöffnet wird.

Die KWIC-Ebene zeigt die Belegzeilen, in welchen ein bestimmter Kollokator zu *cool* (z. B. *Sonnenbrille*) in seiner unmittelbaren Nachbarschaft dargestellt wird. Sie ist ein maximal zeilenlanger Kontextausschnitt, in dem das Stichwort zur Veranschaulichung mittig positioniert wird. Der Vorteil dieser Darstellung liegt in der schnellen Überprüfung des Kontextes sowohl links als auch rechts vom Kollokator. Dabei kann ein kurzer Einblick in die eigentliche Verwendung des Stichworts sowie eine Grobeinschätzung des semantischen Verhältnisses zwischen Lexem und Kollokator gewonnen werden. Die folgenden zwei KWIC-Zeilen demonstrieren die Kontextausschnitte dieser Ebene:

Beispiel *cool* mit Kollokator *Sonnenbrille*

- ┌ B99 nostars dürfen als Werbefiguren für **coole Sonnenbrillen** und Surfboards aufr
- ┌ B00 ganz unschuldig, nicht nur wegen der **coolen Sonnenbrille**. In einer anderen Ze

Vier Vorteile ergeben sich aus dieser Ansicht für die lexikografische Arbeit: Ein Vorteil der KWIC-Ebene liegt in der Möglichkeit, eine schnelle Interpretation der Beziehung zwischen Lexem und Kollokator vornehmen zu können. Sämtliche KWICs eines ausgewiesenen Kollokators erscheinen untereinander, und dieses Ordnungssystem erlaubt ein selektives und systematisches Analysieren von bestimmten Mitspielern und deren Gebrauch. Ein zweiter Vorteil dieser Ebene besteht darin, dass weitere kontextuelle Mit-

¹⁵ In diesem Artikel wird allgemein von Kontext gesprochen, auch wenn es sich in einigen Fällen im strikteren Sinne eher um Kotexte handelt.

¹⁶ Die KWIC-Ebene ist ein Standardformat für Konkordanzen, die immer nur einzeilige Textausschnitte enthält (vgl. Sinclair 1991, S. 43).

spieler auftauchen, die zwar häufig nicht statistisch signifikant sind, aber in Hinblick auf das zu analysierende Lexem Aufschlüsse über semantische Kompatibilität und Wortschatzstrukturen geben. Die Untersuchung der Kontextausschnitte und die damit verbundene Analyse sämtlicher lexikalischer Mitspieler (ob statistisch signifikant oder nicht) muss einen großen Stellenwert für die Lesartenunterscheidung erhalten, da die KWICs Sachgebiete, Referenzbereiche, Domänenbezüge, sprachliche Strukturen und semantisch-syntaktische Verträglichkeiten zwischen Wörtern offen legen. Ein weiterer Nutzen besteht für Lexikografinnen in der Identifikation irrelevanter oder fehlerlematisierter Kollokatoren. Der vierte Vorzug wird darin gesehen, dass der gegebene Kontextausschnitt auch einen Ausschnitt der Prädikation darstellen kann. Der Aussagegehalt der abgebildeten Prädikation ist für eine satzsemantische Klassifikation, wie sie in *lexiko* vorgenommen wird (siehe Abschnitt 3.2), für die theoretische Validierung des Ansatzes einer Lesart von Bedeutung.

Die KWICs sind also als eine Prüfinstanz zu verstehen, welche die Zugehörigkeit der Kollokatoren aus der Kollokationsliste zum gesuchten Stichwort überprüft und deren Relevanz für die Bedeutungsunterscheidung untersucht. Da die KWIC-Ebene meist unvollständige Sätze liefert und in manchen Fällen auch unbrauchbare Zeilenausschnitte anzeigt, kann es vorkommen, dass die extrahierten Informationen der KWIC-Belegzeile nicht ausreichen, um eine vollständige Auswertung eines Kontexts durchzuführen. Hier gilt es, einen entsprechend größeren Sprachausschnitt¹⁷ auf der nächsten Ebene, der des Volltextes, zu analysieren. Die Volltextebene sollte wie die KWIC-Ebene sowohl als Prüfinstanz als auch als Möglichkeit zur zusätzlichen Informationsgewinnung verstanden werden. Aus diesem Grund wird künftig bei der Untersuchung der KWIC- und Volltextebene von Tiefen- oder Verifizierungsanalyse¹⁸ gesprochen. Der Unterschied zwischen den beiden Ebenen besteht darin, dass im Volltext ein entsprechend größeres Informationspotenzial enthalten sein kann, da er den vollständigen semantischen Rahmen eines Lexems und seiner sämtlichen lexikalischen Mitspieler innerhalb eines gewählten Kontextrahmens erfasst. Exemplarisch seien an dieser Stelle die beiden oben ausgewählten KWICs in der Volltextebene angegeben:

Gesellschaftsfähig ist also nur die Idee von Pornografie, die Anspielung auf sie und der Bruch des Tabus. Echte Pornostars dürfen als Werbefiguren für **coole Sonnenbrillen** und Surfboards auftreten, weil die Anspielung dann umso echter wirkt. Kein Musiksender und keine Popband würde je ein Video mit harter Pornografie zeigen. (B99/911.98747 Berliner Zeitung, 13.11.1999, Beilage Magazin; „Dolly is cool“)

In einer Tageszeitung erschien ein Bild von Huber neben Beer, in der Unterzeile wurde der Mann mit der Sonnenbrille als „Bodyguard“ bezeichnet. Huber selbst ist an dem Image nicht ganz unschuldig, nicht nur wegen der **coolen Sonnenbrille**. In einer anderen Zeitung erklärt er, er könne zu dem Attentat auf Beer nichts sagen, weil er nicht zugegen war: „Ich hätte mich sonst vor sie geworfen.“ (B00/006.46705 Berliner Zeitung, 09.06.2000, Politik, S. 6)

Je größer der Volltextumfang eines Lexems ist, desto differenzierter sind die semantischen Angaben, die für die Bedeutungsbeschreibung gewonnen werden können. Die

¹⁷ Die Anzahl der Sätze links und rechts vom Kollokator ist vorher zu selektieren, ist aber aus rechtlichen Gründen auf einen Textausschnitt begrenzt.

¹⁸ Der Begriff der Tiefenanalyse lehnt sich dabei nicht an eine generativistische Terminologie an.

Lexikografin/der Lexikograf analysiert die Regelmäßigkeit des Gebrauchs der Kookkurzen, derjenigen, die in der Kollokationsliste enthalten sind, sowie derjenigen lexikalischen Partner, die als statistisch nicht signifikant im Volltext auftauchen. Auf diese Weise werden Phänomene wie Rekurrenz aufgedeckt, die zusätzliche semantische Aussagen zu sprachlichen Strukturen erlauben. Eine vollständige semantische Analyse zusätzlicher Mitspieler aus der Volltextebene erschließt paradigmatische und syntagmatische Muster, die einen weiteren wesentlichen Beitrag zur Lesartendisambiguierung leisten. Wenn auch essentielle semantische Informationen bereits auf vorangegangenen Ebenen ermittelt wurden, kann nur nach der gründlichen Analyse der Volltextebene eine abschließende Disambiguierung erfolgen. Erst nach einer vollständigen Kontextanalyse sind LexikografInnen in der Lage, linguistische Strukturen satzfunktional zu klassifizieren und damit die vorgenommene Lesartenidentifikation theoretisch zu stützen. Im nächsten Abschnitt wird anhand des Beispielllexems *cool* das traditionelle Verfahren der Disambiguierung, wie es in der Dudenredaktion benutzt wird, mit dem computergestützten Verfahren verglichen.

3. Disambiguierung des Beispielllexems *cool*

In diesem Abschnitt wird das Ziel verfolgt, zwei unterschiedliche Verfahren der Lesartendisambiguierung anhand des Beispielllexems *cool* zu präsentieren. Zunächst wird das Vorgehen der Dudenredaktion beschrieben. Im Anschluss erfolgt die detaillierte Darstellung des softwaregestützten und theoriegeleiteten Verfahrens, wie es in *lexiko* durchgeführt wird. Dabei wird insbesondere auf die Möglichkeiten der korpusbasierten Überprüfung der Disambiguierungskriterien eingegangen.

3.1 Lesarten im Duden

Scholz-Stubenrecht (2001, S. 49; 2002, S. 228) geht auf die Arbeitsabläufe in der Dudenredaktion und auf deren Materialgrundlage ein und hebt dabei die Wichtigkeit der Dudensprachkartei¹⁹ und öffentlich zugänglicher Korpora zur Belegsuche und Stichwortauswahl hervor. Er weist ausdrücklich darauf hin, dass die Korpora zur Belegsuche „für den peripheren Wortschatz“ und „zur statistischen Absicherung lexikografischer Entscheidungen“ herangezogen werden. Die Sprachkartei wird zu Fragen der Stichwortauswahl, der Belegsuche und der Bedeutungsanalyse von Wörtern herangezogen:

Für die Ermittlung neuer Wortbedeutungen oder der Bedeutungen neuer Wörter ist die Duden-Kartei unser zuverlässigstes Hilfsmittel gewesen, da die Belege von menschlichen Exzerptorinnen und Exzerptoren gesammelt werden, die gerade auf diesen Aspekt besonders achten. (Scholz-Stubenrecht 2001, S. 49)

Die aus der Kartei gewonnenen semantischen Informationen werden im Anschluss durch Belege aus Korpora verifiziert. Scholz-Stubenrecht (2001, S. 49) erkennt an, dass „auch Volltextkorpora zumindest ergänzend mit guten Erfolgen genutzt werden“. Obwohl Korpora als geeignetes Arbeitsmittel anerkannt werden, beschränken sich ihre Aufgaben

¹⁹ Die Sprachkartei (früher Kartenkartei, seit 2000 nur noch elektronisch gesammelt) enthält Stichwörter mit Belegen und Quellenangaben und wird durch TextexzerptorInnen ständig ergänzt.

in erster Linie auf die Funktion von Quellenzitaten, um lexikografische Aussagen zu stützen. Hierin besteht der Unterschied zum Korpusansatz und der damit verbundenen *Cosmas*-Methode für das Projekt *elexiko*. Die Arbeitsgrundlage der Dudenredaktion führt für das Beispiellexem *cool* zu folgender Bedeutungsbeschreibung:

cool [ku:l] <Adj.> [engl., eigtl.= kühl] (salopp):

1. *[stets] die Ruhe bewahrend, keine Angst habend, nicht nervös [werdend], sich nicht aus der Fassung bringen lassend; kühl u. lässig, gelassen*: er ist ein extrem -er Typ; als Trainer musst du ziemlich c. sein; lass dich nicht provozieren, bleib [ganz] c.!: Glück muss der Mensch haben, sagte der Kommissar c. (Danella, Hotel 474); es war eben auch nicht so, dass ich das Leben bei meiner Tante auf einmal ganz c. packte (Christiane, Zoo 319).

2. *keine Gefahren bergend, risikolos, sicher*: ein -es Versteck; das ist die -ste Art, den Stoff über die Grenze zu bringen.

3. *keinen, kaum Anlass zur Klage gebend, durchaus annehmbar, in Ordnung*: 1000 Mark ist doch 'n -er Preis für die Anlage.

4. *in hohem Maße gefallen, der Idealvorstellung entsprechend*: auf der Fete waren unheimlich -e Leute; die Musik war echt c.; New York ist die -ste Stadt überhaupt (Popcorn 10, 1988, S. 7); die Band spielt irrsinnig c.

© 2000 Dudenverlag

Die Einteilung in vier Lesarten scheint zunächst nachvollziehbar, auch wenn der Beleg für Lesart 2 einen großen Interpretationsspielraum besitzt und die zugewiesene Bedeutung nicht eindeutig nachweisbar ist. Die semantische Information des Dudenartikels beschränkt sich neben der Anzahl der Lesarten auf die Bedeutungsparaphrasierung inklusive Angaben zu Synonymie, die aber nicht explizit ausgewiesen sind, und Belegbeispielen. Eine detaillierte Auskunft zu Verwendungscharakteristika oder Angaben zu den Beziehungen zwischen den einzelnen Lesarten enthält der Eintrag nicht. Es fehlen außerdem thematische Angaben zu Sachgebieten und pragmatische Angaben zum Bewertungspotenzial des Lexems. Dieser Mangel ist aber durch die Platzbeschränkung eines Printwörterbuchs zu rechtfertigen. Es bleibt festzuhalten, dass keine klaren Prinzipien der Disambiguierung eines Lexems im Duden GWDS (2000) erkennbar sind.

3.2 Lesarten in *elexiko*

Auf der Grundlage der Korpora, der automatischen Kookkurrenzanalyse, der Verifizierungsanalysen und einer kontrollierten vernetzten Kriteriologie versucht *elexiko* einer theoriegeleiteten aber praktikablen Lesartendisambiguierung näher zu kommen. Die Kriteriologie, wie sie für das Projekt *elexiko* herausgearbeitet worden ist, bildet die Grundlage dafür, die kontextuellen Identifikationen und Differenzierungen der Lesarten eines Lexems nach ihrer Funktion entweder satzsemantisch oder syntaktisch zu klassifizieren. Die funktionale Betrachtung des Beitrags eines Lexems zur Satzbedeutung steht dabei im Vordergrund. Die Einteilung der Lesarten nach funktional-propositionalen Kri-

terien bietet den Vorteil, dass diese das Lexem in seinen unterschiedlichen Verwendungsweisen und semantischen/syntaktischen Funktionen abbildet. Die Vorteile der Lexembeschreibung nach funktionalen Wortklassen anstatt nach traditionellen Wortarten im Wörterbuch sieht Strauß (1989, S. 788-796) in der Darstellung des Zusammenhangs zwischen semantischer Form und Proposition bzw. illokutivem Potenzial. Auch Pustejovsky (1995, S. 62) verweist auf die Notwendigkeit der Beschreibung der Funktionen eines Lexems, um es als integratives Element der Satzbedeutung zu erfassen:

By defining the functional behaviour of lexical items at different levels of representation we hope to arrive at a characterization of the lexicon as an active and integral component in the composition of sentence meanings.

Die zugrunde liegenden Kriterien wurden in Anlehnung an die satzsemantische Klassifikation von v. Polenz (1988, S. 159-166) für die lexikografischen Bedürfnisse des Projektes *lexiko* weiterentwickelt. Die Disambiguierung des Adjektivs *cool* stützt sich auf folgende semantische Kriterien²⁰, deren Prüfung auf den unterschiedlichen Ebenen empirisch vorgenommen wird:

- satzsemantische Klassifikation
- paradigmatische Muster
- Gruppierung von Kookkurrenzen nach Nomen und deren Referenz
- syntagmatische Muster, insbesondere Gebrauch von Adjektivphrasen mit Modifikatoren.²¹

Sowohl paradigmatische als auch syntagmatische Muster sind klassische Disambiguierungskriterien der strukturellen lexikalischen Semantik.²² Diese werden darüber hinaus in *lexiko* um eine satzfunktionale Klassifikation, um eine referenzielle Klassifikation (hier Gruppierungen nominaler Kookkurrenzen nach ihrer Referenz) sowie um eine Gruppierung syntagmatischer Muster (hier Gebrauch von Modifikatoren) bereichert. Reichmann (1989) betont die Leistungsfähigkeit, die durch das Zusammenspiel der verschiedenen Kriterien entsteht.²³ Mithilfe dieser Kriterien wird der unterschiedliche Gebrauch (also Funktionen und Verwendungsmuster) des Lexems *cool* gruppiert.²⁴ Die Einteilung der Verwendungsmuster in unterschiedliche linguistische Klassen bildet die Grundlage für das Ansetzen separater Lesarten. Nach vollständiger Datenanalyse und -interpretation zeigt sich, dass für das Adjektiv *cool* vier Kategorien satzsemantischer Funktionen²⁵ definiert werden können: Das Lexem *cool* tritt als Zustandsprädikator, Ei-

²⁰ Die Wahl der Disambiguierungskriterien hängt stark mit dem zu untersuchenden Lexem *cool* zusammen, da es sich hier um ein Adjektiv handelt. Grundsätzlich gilt, dass einige Kriterien bei einer anderen Wortart nicht disambiguierend fungieren und daher weitere Kriterien hinzugefügt werden müssen. Während z. B. bei Adjektiven das Kriterium der Argumentrollen unwichtig ist, wird dieses Kriterium bei Verben eine entscheidende Position einnehmen. Die zugrunde liegenden Kriterien spiegeln z.T. bedeutungskonstituierende Strukturen im Sinne Pustejovskys (1995) wider. Vgl. Inhalte der *qualia structure*, *event structure*, *argument structure* und *lexical inheritance structure*.

²¹ Als Modifikatoren treten Adverbien und Intensitätspartikeln auf.

²² Vgl. Disambiguierungskriterien bei Reichmann (1989, S. 111-114).

²³ Vgl. Pustejovskys (1995) Vorstellung der vier Strukturelemente, die z.T. in den Disambiguierungskriterien abgebildet sind.

²⁴ Vgl. auch Sinclairs (1987) 'use and function of a word'.

²⁵ Diese Kategorien sind vergleichbar mit Pustejovskys (1995) *lexical conceptual paradigms*, die semantisch unterspezifiziert sind und erst im Kontext ihre semantische Spezifikation erhalten.

genschaftsprädikator, Einstellungsprädikator und als Referenzwort²⁶ auf. Darüber hinaus enthalten diese vier satzsemantischen Formen unterschiedliche Lesarten, die über ihre paradigmatischen, syntagmatischen, referenziellen oder inhaltlich-thematischen Kollokatoren voneinander abgrenzbar sind. Besonders Nomen, die als Kollokatoren auftreten, werden durch das Adjektiv *cool* modifiziert. Die Referenz der modifizierten Nomen spielt dabei eine entscheidende Rolle für die Disambiguierung. Eine Hierarchisierung der einzelnen Lesarten wird an dieser Stelle nicht vorgenommen.²⁷ Die Markierung der Lesart erfolgt üblicherweise verbal in einfachen Anführungszeichen. Zur kurzen Identifikation der Lesarten dient eine Etikettierung durch Synonyme, wobei die Klasse der Referenzwörter dabei eine Ausnahme bildet.

Satzsemantische Funktion: Zustandsprädikator

- Lesart: 'kalt, kühl' (Das Adjektiv *cool* modifiziert Nomen, mit Referenz: Temperaturzustand)
- Lesart: 'ruhig' (Das Adjektiv *cool* modifiziert Nomen, mit Referenz: Gemütszustand)

Satzsemantische Funktion: Einstellungsprädikator

- Lesart: 'leicht abweisend, lässig, ignorant' (Das Adjektiv *cool* modifiziert Nomen, die auf positive oder negative Einstellungen oder auf das Bewerten von (meist menschlichem) Verhalten referieren.)
- Lesart: 'akzeptabel, in Ordnung' (Das Adjektiv *cool* modifiziert Nomen, die auf positive Einstellungen oder auf das Bewerten von Eigenschaften, Zuständen oder Dispositionen referieren.)
- Lesart: 'super, sehr gut gefallen, interessant' (Das Adjektiv *cool* modifiziert Nomen, die auf sehr positive Einstellungen oder auf das Bewerten von Eigenschaften, Zuständen oder Dispositionen von etwas (meist Artefakten) referieren.)

Satzsemantische Funktion: Eigenschaftsprädikator

- Lesart: 'einfach, schlicht, nüchtern' (Das Adjektiv *cool* modifiziert Nomen, die auf Eigenschaften oder Dispositionen einer Sache referieren.)
- Lesart: 'modern, aktuell' (Das Adjektiv *cool* modifiziert Nomen, die auf Eigenschaften oder Dispositionen einer Sache mit Zeitbezug referieren.)

Satzsemantische Funktion: Referenzwort

- Referenz: Personen
- Referenz: Musikrichtung
- Referenz: Film/Theaterstück etc.
- Referenz: Institution
- Referenz: Produkt

²⁶ Referenzwort wird hier im weitesten Sinne als Eigenname verstanden.

²⁷ Im Korpus ist aber deutlich zu erkennen, dass die individuellen Lesarten unterschiedlich stark verteilt sind. Die Frage, ob die frequentesten Lesarten auch die prototypischsten oder allgemeinsten Lesarten sind, soll hier nicht diskutiert werden. Der Fokus dieser Arbeit liegt ausschließlich auf der Beschreibung des Disambiguierungsverfahrens. Zu Prinzipien und Problemen der Lesartenanordnung in allgemeinen einsprachigen Wörterbüchern vgl. Werner (1989, S. 917-930).

3.2.1 Zustandsprädikatoren

Zu der satzsemantischen Kategorie 'Zustandsprädikator' zählen Wörter, mit denen Aussagen über einen Zustand oder Status von Lebewesen, Sachen oder Abstraktbegriffen getroffen werden. Zustände sind dadurch gekennzeichnet, dass sie zu einem Zeitpunkt eintreten und enden, damit also grundsätzlich veränderbar sind.

3.2.1.1 'kalt, kühl'

Cool in der Lesart 'kalt, kühl' referiert auf einen Temperaturzustand. Obwohl diese Lesart im Vergleich zu den anderen Lesarten des Lexems nicht sehr frequent ist, existieren drei Indikatoren²⁸ in der Kollokationsliste, die auf diese Lesart aufmerksam machen:

BelegNr	Gamma	Kollokatoren	Häufigkeit
2190+3:	579	Drink heiss bar	3
2193+12:		Drink heiss	12
2205+5:		Drink bar	5
2210+36:		Drink	36

Während der erste Indikator *Drink* den referenziell-thematischen Anwendungsbereich des Temperaturzustandes auf Getränke markiert, fällt der Kollokationspartner *heiß* durch sein antonymisches Verhalten auf. Der dritte Indikator *Bar*²⁹ beschreibt eine Domäne im weitesten Sinne und steckt damit einen prototypischen Gebrauchsrahmen für den vorangegangenen Kollokator *Drink* ab. Das folgende Beispiel illustriert die Lesart anhand seiner semantischen Nachbarschaft zu *heiß* und *Drink*.

Beispiel:

Natürlich werden bei der Havanna-Club-Beach-Party auch karibische und südamerikanische Klänge im Vordergrund stehen. In den Umbauphasen sorgt DJ K.A.I für die passenden Töne, damit die Stimmung nicht abreißt. Heiße Musik, heiße Tänze, heiße Temperaturen und coole Drinks, alles für die Renovierung des maroden und in die Jahre gekommenen Waldstadions, das gerade seinen 40. Geburtstag feiert. Im Terminkalender der Viernheimer Musik- und Partyfreunde ist der 19. August (ab 18 Uhr) sicher schon vermerkt. In der kommenden Woche beginnt der Vorverkauf. (M00/007.40175 Mannheim Morgen, 22.07.2000, Lokales; Beach-Party steigt am 19. August)

Untersucht man auf tieferen Analyseebenen die Textbeispiele, ergeben sich weitere paradigmatische Partner, die nicht mittels Kookkurrenzliste identifiziert werden konnten.³⁰ Aus den Kookkurrenzbelegen können folgende Angaben ermittelt werden: Synonym – *kühl*, Antonym – *heiß*, Kohyponym – *erfrischend*. Da diese Lexeme in die gleiche satzsemantische Gruppe der Zustände mit Referenz auf einem Temperaturzustand klassifiziert werden können, demonstrieren sie in ihren Kontexten ihren direkten paradigmatischen

²⁸ Die Kollokatoren sind absteigend nach Stärke der lexikalischen Kohäsion sortiert. Sie wurden hier übersichtshalber zusammengefasst, in der Kollokationsliste erscheinen diese aber nicht untereinander.

²⁹ In der Kollokationsliste ist *bar* kleingeschrieben, was auf einen Lemmatisierungsfehler zurückzuführen ist. Hier wurden *bar* und *Bar* dem gleichen Lemma zugeordnet.

³⁰ Die Gründe, warum diese nicht in der Kookkurrenzliste auftauchen, sind vielschichtig und wiederum von den *Cosmas*-Einstellungsoptionen abhängig. Zum einen kann es sich um die Auswahl der Stichproben des Zufallsgenerators handeln, zum anderen um den okkasionellen Gebrauch, dem keine weitere usuelle oder statistische Signifikanz zugeschrieben wird.

schen Bezug zu *cool*. Durch die Untersuchung sämtlicher KWICs auf der Volltextebene wird deutlich, dass der Zustand oft Flüssigkeiten zugeschrieben wird. Verdeutlicht wird dieser Markierungsbereich ebenfalls durch die Mitspieler *Wasser*, *Getränk*, *Schwimmbad*, deren Auftreten aber nicht statistisch signifikant ist. Sie bekräftigen den Anwendungsbereich für *cool* als Modifikator für die Beschreibung eines Temperaturzustandes und validieren das Ansetzen dieser Lesart.

In einigen Kontexten ist es auffällig, dass diese Lesart ambig ist und Doppeldeutungen erkennbar sind, indem bei ein- und derselben Textstelle sowohl die Interpretation 'kalt, kühl' als auch die Interpretation 'super, sehr gut gefallend, interessant' zugelassen ist. Folgendes Beispiel demonstriert dieses Verhältnis zwischen den beiden Lesarten:

Beispiel:

„Coole Drinks“ erfrischen in der „Waikiki-Bar“ des jungen DLRG-Teams, gegenüber parkte eine Phalanx heißer Öfen vor dem Stand des Biker-Clubs. Und vis-à-vis hatte das Rote Kreuz sein Bereitschaftszelt aufgeschlagen. Sogar die Youngsters vom Edinger Jugendrotkreuz opferten hier ehrenamtlich ihre Freizeit. (M99/907.45745 Mannheimer Morgen, 14.07.1999, Lokales; Hanseaten schmetterten Seemannslieder)

Nicht selten kann sich bei solchen Beispielen die Konstruktion *coole Drinks* sowohl auf die Temperatur als auch auf die exotische Zusammenstellung eines Getränks, die als interessant und wohlgefällig bewertet wird, beziehen. SprecherInnen und SchreiberInnen machen sich in vielen Kontexten das ambige Verhältnis des Lexems daher auch zunutze.

Der Dudenwortartikel enthält diese Lesart nicht. Die Ursachen dafür könnten in der Methode der Bedeutungsanalyse liegen. Da in der Dudenredaktion Korpora nicht zur Erkenntnisgewinnung, sondern hauptsächlich zur Belegsuche genutzt werden, und somit keine systematisch vorstrukturierte Datenanalyse eines sehr großen Datenbestandes durchgeführt wird, wurde diese Ambiguität nicht erfasst.

3.2.1.2 'ruhig, gelassen'

Eine weitere Zustandsbeschreibung des Lexems *cool* wird in der Lesart 'ruhig, gelassen (die Ruhe bewahrend)' realisiert. Diese Lesart charakterisiert Beschreibungen eines Gemütszustands, insbesondere von Menschen. Folgende Kollokatoren indizieren den Gebrauch des Lexems *cool* in dieser Lesart:

BelegNr	Gamma	Kollokatoren	Häufigkeit
1062+50:	1314	ganz bleiben	50
1112+356:		ganz	356
1530+25:	1125	bleiben immer	25
1555+428:		bleiben	428
3228+1483:	295	sein	1483
5891+3:	82	äusserlich innerlich	3
5894+1:		äusserlich	1
5132+2:	61	innerlich	2
5858+14:	68	völlig	14

Eine Verifizierung der identifizierten Lesart ist nur durch die Tiefenanalyse möglich, wie die folgenden Beispiele illustrieren.

„Wir wollen bei einem Sicherheitslehrgang keinen Crash riskieren.“ Und der folgt oft, wenn ein Bus ins Bankett fährt. „Das läßt sich häufig vermeiden“, erklärt der Sicherheitstrainer, „wenn es dem Fahrer gelingt, ganz cool zu bleiben, einfach geradeaus zu fahren und sein Fahrzeug so zu stabilisieren.“ Wobei „einfach“ überhaupt nicht einfach ist. Aber es hilft ungemein, zu sehen, in welcher vermeintlich aussichtsloser Situation noch fahrerische Rettung möglich ist.
(FAZ/548.00048: Frankfurter Allgemeine Zeitung, 1995)

Zunächst unterlag ein tobender und wild mit seiner Unfähigkeit hadernder Boris Becker („Ich treffe einfach gar nichts, raus hier, raus, raus!!!“) dem völlig cool und fehlerfrei daher spielenden Andre Agassi 5:7, 3:6, 6:3, 1:6 und vertat damit die Chance, Nummer eins zu werden. (T91/JUN.23143 die tagesschau, 10.06.1991, S. 17; Monica Seles trotzte „Big Brother“)

Diese Lesart ist im Duden GWDS (2000) (siehe Lesart 1) enthalten, wird aber allgemeiner verstanden. Im Gegensatz zur Dudenlesart 1 (‘[stets] die Ruhe bewahrend, keine Angst habend, nicht nervös [werdend], sich nicht aus der Fassung bringen lassend; kühl u. lässig, gelassen’) werden die Ergebnisse der *Cosmas*-Recherche nach unterschiedlichen Kriterien in zwei eigenständige Lesarten (‘ruhig, gelassen’ und ‘leicht abweisend, lässig, ignorant’) unterteilt. Eine Disambiguierung nach dem Kriterium des unterschiedlichen Gebrauchs von Modifikatoren greift in diesem Zusammenhang nicht, da die Modifikatoren *völlig* und *ganz* bei der Verwendung beider Lesarten genutzt werden. Das würde dafür sprechen, beide Lesarten als eine Lesart, wie sie im Duden GWDS (2000) vorgenommen wurde, zu verstehen. Auffällig ist aber, dass beide Lesarten in unterschiedliche satzsemantische Kategorien eingeordnet werden. Während es sich bei Lesart ‘ruhig, gelassen’ um einen Prädiktor des Zustands handelt, handelt es sich bei Lesart ‘leicht abweisend, lässig, ignorant’ um eine Prädikation der Einstellung. Durch die Verifizierungsanalyse ist zu erkennen, dass die paradigmatischen Kookkurrenzen auch entsprechend zwei individuelle semantische Muster aufbauen, die auf zwei unterschiedliche kontextuelle Verwendungsweisen hindeuten. Demnach wirken das Kriterium der satzsemantischen Klasse und das Kriterium der paradigmatischen Muster lesartenunterscheidend.

Besonders im Bestreben, unterschiedliche paradigmatische Muster³¹ zu erkennen, sollte man sein Augenmerk auf die Analyse tieferer Ebenen richten. Auf diese Weise konnten folgende Synonyme ermittelt werden: *entspannt, kontrolliert, laid back, relaxed, ruhig, zurückgelehnt*. Diese lexikalischen Mitspieler können nicht als Synonyme für die Lesart ‘leicht abweisend, lässig, ignorant’ eingesetzt werden. Eine Reihe weiterer interessanter Bindeglieder (*beherrscht, easy, freundlich, lonelyhearted, selbstsicher, traurig, trocken, zurückhaltend*) taucht in Kontexten auf, deren eindeutige Eingruppierung in synonyme oder kohyponyme Felder nicht ohne Schwierigkeiten erfolgen kann. Hier sollte auf eine detaillierte Trennung verzichtet werden, und stattdessen sollten diese Mitspieler in einem Feld, das aus koreferenziellen Partnern mit gemeinsamen Referenzbezug besteht,

³¹ Das paradigmatische Feld beinhaltet viele Kookkurrenzen, die u.a. der Jugendsprache angehören. Das wird anhand der Anglizismen deutlich. Wie diese wiederum lexikografisch eingebettet werden können, soll an dieser Stelle nicht erörtert werden.

zusammengefasst werden. Antonyme können in inkompatiblen Propositionen auftreten, auf die durch inkompatible Paarwörter wie die Folgenden aufmerksam gemacht wird:

BelegNr	Gamma	Kollokatoren	Häufigkeit
5891+3:	82	äusserlich innerlich	3
5894+1:		äusserlich	1
5132+2:	61	innerlich	2

Aus den Textbelegen des Mitspielers *innerlich* gehen folgende Antonyme hervor: *be-bend, erregt, kochend, sturmreif, tobend, wild*. Auch bei den Antonymen handelt es sich jeweils um graduierbare Adjektive, die semantisch eine Zustandsbeschreibung im Sinne eines Gemütszustandes definieren. Die gezielte introspektive Suche nach Antonymen seitens der Lexikografinnen ergibt folgende Ergänzungen: *nervös, hektisch, cholerisch*. Diese sind jedoch nicht in der Kookkurrenzliste oder auf der KWIC- bzw. Volltextebene zu finden, da sie aufgrund geringer Frequenz vom *Cosmas*-Rechercheprogramm keine statistische Signifikanz zugeschrieben bekommen. Hier können nur Belege extrahiert werden, die durch gezielte introspektive Rückprüfung ermittelt werden konnten, um die vorgenommene Lesartendisambiguierung zu stützen.

3.2.2 Einstellungsprädikatoren

Mittels Einstellungsprädikatoren können positive und/oder negative Bewertungen eines Zustands oder einer Eigenschaft von Personen bzw. Objekten vorgenommen werden. Besonders Schimpf-, Schlagwörter sowie Hochwertwörter, die mit entsprechenden pragmatischen Angaben in Verbindung gebracht werden, zählen zu dieser Gruppe. Nach diesem Kriterium sind für das Lexem *cool* folgende Lesarten als Einstellungsprädikatoren anzusetzen:

- 'leicht abweisend, lässig, ignorant': impliziert sowohl positive als auch negative Bewertung
- 'akzeptabel, in Ordnung': impliziert positive Bewertung
- 'super, sehr gut gefallend, interessant': impliziert extrem positive Bewertung

Die Differenzierung der drei Lesarten wird mit den unterschiedlichen paradigmatischen Verwendungsweisen und dem unterschiedlichen Gebrauch von Modifikatoren begründet, wie im Anschluss detailliert dargestellt wird.

3.2.2.1 'leicht abweisend, lässig, ignorant'

Die Lesart 'leicht abweisend, lässig, ignorant' ist im Verhältnis zu anderen Lesarten des Lexems *cool* sehr frequent im Korpus belegt. Sie kommt insbesondere in Kontexten vor, in denen die Beschreibung und Bewertung einer Person im Vordergrund steht. Vermutlich ist diese Lesart mit der Lesart 'kalt, kühl' verwandt und durch metaphorische Konzeptverschiebung entstanden. Dabei könnte der Temperaturzustand 'kalt, kühl' auf menschliches Verhalten im Sinne von 'kühl, ohne Gefühle zu zeigen' übertragen worden sein. Zuschreibungsobjekte wie Personen und Objekte, die diese Bewertung erfahren, finden in der Kollokationsliste deutlich ihren Niederschlag:

BelegNr	Gamma	Kollokatoren	Häufigkeit
740+169:	1381	Typ	169
2251+90:	658	Kid	90
2928+63:	344	Jung	63
5222+17:	136	Girl	17
5430+5:	117	Rapper	5
5681+8:	100	Boy	8
5954+4:	72	Gangster	4

Die Ergebnisse der KWIC-Auswertung zeigen, dass aufgrund semantischer Überlappungen nicht alle Kollokatoren eindeutig einer bestimmten Lesart zuzuordnen sind. In diesem Fall können die Lesarten nicht eindeutig unterschieden werden, da ein großer Spielraum für unterschiedliche Interpretationen existiert. Betrachtet man nun auf der Volltextebene einen Kollokator, so kann geprüft werden, inwieweit sich diese Kollokatoren als Mitspieler der gewünschten Lesart verifizieren lassen. Das folgende Beispiel mit den Mitspielern *lässig*, *keine Skrupel* soll diesen Verifizierungsprozess demonstrieren.

Ganz cooler Typ, der *lässig* sein Tuch um den Hals geschwungen trägt, einer, den *keine Skrupel* plagen, wenn es darum geht, den Barbaren das Bildnis der Göttin Diana abzugeben. (M89/901.03419: Mannheimer Morgen, 28.01.1989)

Bei der Untersuchung der einzelnen substantivischen Mitspieler aus der Kollokationsliste und derjenigen, die aus der Tiefenanalyse hervorgegangen sind (*Single-Frau*, *Star*, *Mädchen*, *Frau*), wird ein bestimmtes semantisches Feld für Personenzuschreibungen erkennbar. Darüber hinaus kann ein weiteres Feld durch verbale Mitspieler, die sich auf das Bewerten mittels *cool* beziehen, etabliert werden.

BelegNr	Gamma	Kollokatoren	Häufigkeit
5586+53:	107	wirken	53
5757+8:	92	rauchen	8
6138+3:	59	gucken	3
6318+12:	54	auftreten	12

Diese Gruppierung kann auch durch die folgenden Kollokatoren erweitert werden, die sich unter der Fragestellung 'was gilt als cool?' zusammenfassen lassen: *auftreten*, *Gang*, *Großschnäuzigkeit*, *Image*, *leugnen*, *Macho*, *Mimik*, *Pokerface*, *rauchen*, *stehen*, *wirken*, *Zigarette*. Einige der Mitspieler sind bereits der Kookkurrenzliste zu entnehmen, andere können erst durch die Volltextuntersuchung extrahiert werden.

Besonders markant kommt das Bewertungspotenzial bei paradigmatischen Kookkurrenzen zum Ausdruck, wobei eine exakte Trennung zwischen Synonymie und Hyponymie nur schwer möglich ist. Da beide Sinnrelationen eng miteinander verwandt sind (die Synonymie ist ein Spezialfall der Hyponymie), sollten sie als koreferenzielle Partner zusammengefasst werden. Folgende koreferenzielle Kookkurrenzen bieten einen semantischen und pragmatischen Einblick in den Bedeutungsumfang dieser Lesart:

BelegNr	Gamma	Kollokatoren	Häufigkeit
2551+50:	412	lässig	50
5965+6:	70	unnahbar	6
6237+8:	58	frech	8
6304+5:	56	abgeklärt	5

Die Rekurrenz, auf die man in der Verifizierungsanalyse immer wieder stößt, erschließt in manchen Fällen sehr komplexe paradigmatische Strukturen. So werden neben den oben genannten paradigmatischen Partnern aus der Kollokationsliste folgende bedeutungserschließende Kookkurrenzen zur Lexembeschreibung gewonnen:

abgebrüht, abgeklärt, aggressiv, böse, clever, dreist, durchtrainiert, frech, hart, individuell, kalkulierend, karriere-orientiert, lässig, locker, männlich, perfekt, provozierend, salopp, schön eigen, schön faul, schräg, selbstbewusst, sportlich, stark, tough, überlegen, unnahbar, verwegen, wortkarg, zornig, zynisch

Die Mehrheit dieser Lexeme ist nicht statistisch signifikant, aber die Komplexität dieses paradigmatischen Netzes veranschaulicht beispielhaft den semantischen Umfang des Lexems *cool* in dieser Lesart. Während die Belege von Koreferenz sehr zahlreich sind, sind Antonyme (*uncool, sinnlich, weich, romantisch*) selten präsent. Auffällig ist dabei die semantische Flexibilität der negierten Form *uncool*, die auch durch die unnegierte Form *cool* zutage tritt. Die Paradigmatik kristallisiert sich aufgrund ihrer informativen Komplexität auch bei den Einstellungsprädikaten von *cool* als ein essentielles Abgrenzungskriterium der Lesarten heraus.

Modifikatoren zu *cool*, die in Verbindung mit dieser Lesart auftauchen, sind: *extrem, richtig*. Sie grenzen jedoch nicht die Lesarten innerhalb der Kategorie Einstellungen/Bewertungen voneinander ab, da auch eine weitere Lesart 'super, sehr gut gefallend, interessant' mit diesen Modifikatoren zusammen auftritt. Sie grenzen aber die beiden genannten Lesarten von der dritten Lesart 'akzeptabel, in Ordnung' der gleichen satzsemantischen Klasse ab – der einzigen Lesart dieser Klasse – in deren Gebrauch *cool* nicht mit *extrem, richtig* modifiziert werden kann. Die Hauptaufgabe dieser Modifikatoren liegt daher in der Disambiguierung zwischen den satzsemantischen Klassen, denn das Lexem *cool* taucht in anderen Prädikationsfunktionen nicht in dieser syntagmatischen Nachbarschaft auf.

3.2.2.2 'akzeptabel, in Ordnung'

Eine als positiv bewertete Einstellung spiegelt sich in den Kontexten wider, die mit 'akzeptabel, in Ordnung' etikettiert werden können – eine Lesart, die auch im Duden GWDS (2000) genannt ist. Folgende Kookkurrenzen kristallisieren sich dabei als Indikatoren heraus:

BelegNr	Gamma	Kollokatoren	Häufigkeit
2481+70:	469	ziemlich	70
4907+16:	205	passabel	16
5435+6:	117	3=cool	6
5567+6:	100	1=schwach	6

Bei der Tiefenanalyse findet man einen interessanten Beleg, der auf eine Bewertungsskala Bezug nimmt, in der *cool* als zweitbeste Bewertung eingestuft wird:

CD-Tip: 1 = schwach, 2 = passabel, 3 = **cool**, 4 = hervorragend Die New Yorker «Fun Lovin' Criminals» sind einfach cool! Angefangen beim coolen Drumsound, wo schon die erste Nummer den Tarif durchgibt. Weiter mit den coolen Bassloops: Die sind Rhythmus pur, manchmal durch einen einzigen Ton. (A98/OKT.685386 St. Galler Tagblatt, 02.10.1998)

Der Gebrauch der Bewertung von *cool* ist klar von den anderen Wertungen abgegrenzt. Während es sich bei dieser Kollokatorenliste (*schwach*, *passabel*, *hervorragend*) um Kohyponyme handelt, werden bei der Verifizierungsanalyse weitere paradigmatische Mitspieler, die im weitesten Sinne synonymische sind (*gut*, *nicht übel*), ermittelt. Da die Bewertung in dieser Lesart als positiv empfunden wird, ist sie nicht leicht von der im Anschluss beschriebenen Lesart 'super, sehr gut gefallend, interessant', bei der *cool* als bestbewertete Einstellung hervorgeht, zu unterscheiden, stehen doch beide Lesarten in einer inkompatiblen Relation zueinander. Problematisch ist auch das Dudenbeispiel *1000 Mark ist doch 'n -er Preis für die Anlage*. Hier kann keine eindeutige Bewertung im Sinne der Lesart 'akzeptabel, in Ordnung' im Kontrast zu 'super, sehr gut gefallend, interessant' erfolgen. Dieser Beleg kann beiden Lesarten zugeordnet werden. Es hätte also ein treffenderer Beleg gefunden werden müssen. Die systematische Analyse der Kookkurrenzen und ihrer Kontexte bietet in diesem Fall einen Vorteil gegenüber der Introspektion und Belegsuche durch Rückprüfung in den Korpora. Das Problem der Belegauswahl beschreibt Sinclair (1991, S. 4) wie folgt:

Especially in lexicography, there is a marked contrast between the data collected by computer and that collected by human readers exercising their judgment on what should or should not be selected for inclusion in a dictionary.

Die Gesamtkookkurrenzliste enthält nur wenige Indikatoren, die auf die Lesart 'akzeptabel, in Ordnung' aufmerksam machen. Sowohl die Anzahl der Belege als auch die Ergebnisse der Tiefenanalyse zeigten, dass diese Lesart nicht frequent belegt ist. Die dominierende Präsenz der Lesart 'super, sehr gut gefallend, interessant' in den Korpora erschwert eine Lesartendisambiguierung. Sinclair (1991, S. 55) weist in seinem Disambiguierungsbeispiel des englischen Lexems *yield* auf ein ähnliches Phänomen hin, für das er die Disambiguierungsschwierigkeiten zwischen „starken“ und „schwachen“ Lesarten beschreibt:

It does not actually have a strong meaning, and so it does not contrast sharply with the others.

3.2.2.3 'super, sehr gut gefallend, interessant'

Diese Lesart ist zusammen mit der Lesart 'leicht abweisend, lässig, ignorant' am häufigsten im Korpus vertreten. Damit stellt sich *cool* mit seinem Bewertungspotential dominant vor andere mögliche Bedeutungsinterpretationen, die Vielzahl an Indikatoren überrascht daher nicht. Die positive Einstellung, die in dieser Lesart beschrieben wird, bezieht sich auf die Bewertung von Überdurchschnittlichem.

BelegNr	Gamma	Kollokatoren	Häufigkeit
909+49:	1328	echt sein	49
958+104:		echt	104
2722+46:	363	geil	46
2768+52:	354	total	52
2820+47:	350	einfach sein	47
2867+61:		einfach	61
3222+1:	309	stilvoll	1
2991+231:	331	finden	231
3299+10:	298	voll Angus	10
3309+60:		voll	60
5157+23:	151	wirklich	23
5252+40:	133	richtig	40
5348+49:	124	sehr	49
5410+16:	117	absolut	16
5467+11:	114	sexy	11
5912+18:	76	besonders	18
6144+4:	59	unheimlich	4
6289+3:	57	megageil	3

Bei diesen Indikatoren fällt auf, dass verschiedene syntagmatische Partner, die *cool* spezifizieren, vertreten sind. Wie der Inhalt der Lesart selbst, tragen einige der Modifikatoren, wie z. B. *total*, *voll*, *absolut*, Superlativcharakter. Diese können als intensivierende Modifikatoren verstanden werden, deren Bedeutungsinhalt eine Abgrenzung zur Lesart ‚akzeptabel, in Ordnung‘ begünstigt und die inhärenten Merkmale von *cool* bekräftigend zum Ausdruck bringt, wie das folgende Beispiel besonders gut verdeutlicht:

Langweilig, das ist für die zwölfjährigen Schülerinnen ihr Zuhause in Berlins Trabantenstadt Marzahn, „wo es nur Hochhäuser und Kaufhallen gibt“. Janina findet den Potsdamer Platz sogar „cool, echt cool“. Vor allem das Straßengeviert rund um das CinemaxX. (T98/OKT.43892 die tageszeitung, 02.10.1998, S. 3)

Die Liste der Mitspieler enthält auch Ausdrücke der Koreferenz, z. B. *geil*, *megageil*, *sexy*, *stilvoll*. Unterstützende Bewertungsfunktion erhält das Lexem auch durch den verbalen Mitspieler *finden*. Bei den substantivischen Kookkurrenzen ist eine klare Zuordnung zu dieser Lesart nicht ohne Volltextrecherche möglich. Diese Mitspieler leisten bei bloßer Betrachtung der Kookkurrenzliste keine klare Lesartenidentifikation.

BelegNr	Gamma	Kollokatoren	Häufigkeit
3369+1:	296	Session	1
5300+6:	128	Party	6
5329+18:	127	Disco	18
6134+4:	61	Location	4

Bei diesen Kollokatoren ist zu prüfen, ob in der Prädikation die Beschreibung einer Eigenschaft oder eine Bewertung im Vordergrund steht. So können viele Mitspieler auch als Indikatoren der Lesart ‚modern, aktuell‘ interpretiert werden.

BelegNr	Gamma	Kollokatoren	Häufigkeit
2676+46:	368	Sonnenbrille	46
5770+6:	90	Klamotte	6
5933+9:	75	Beat	9
6036+8:	67	Design	8
6044+24:	67	Musik	24
6106+7:	65	Frisur	7
6148+1:	59	Soundsystem	1

Diesem Interpretationsfreiraum sind auch andere Kollokatoren unterworfen:

BelegNr	Gamma	Kollokatoren	Häufigkeit
5669+4:	103	bewegen	4
5709+19:	96	aussehen	19

Nur durch die Volltextauswertung gelingt es, die Interpretationslücken bis zu einem gewissen Maß zu schließen.

Pragmatisch interessant sind die Themengebiete Musik und Kleidung, die sich durch eine Gruppierung der substantivischen Kollokatoren (*Party, Disco, Musik, Song, etc.*) erschließen lassen. Solche inhaltlich-thematischen Felder drücken die Anwendungsbereiche bzw. Domänen für die Verwendung des Lexems in dieser Lesart aus.

Ein selbstsprechendes Bewertungsfeld semantisch eng verwandter Lexeme lässt sich aus der KWIC-Ebene extrahieren. In den meisten antonymischen Kontexten fungiert *uncool* als semantisch flexible Negierung.³² Dagegen gibt es eine wahre Flut an synonymen oder kohyponymen Mitspielern, die in ihrer Gesamtheit eine Paraphrasierung des Lexems *cool* darstellen. Einige dieser kontextuellen Mitspieler agieren rekurrent, tauchen nur vereinzelt in Belegen auf, andere mehrfach. Das umfassende koreferenzielle Feld veranschaulicht die Abgrenzung von der Lesart 'modern, aktuell', deren paradigmatische Netze anders ausgeprägt sind:

anziehend, außergewöhnlich, beautiful, beliebt, besser, crazy, einfach genial, einmalig, exzellent, exklusiv, edel, erfolgreich, geil, grandios, mega, megageil, high, innovativ, Klasse, kult, kultiviert, originell, reizvoll, schickschön, schön, schwungvoll, sexy, spitze, spritzig, super, superb, unschlagbar, unvergleichbar, unverwechselbar, vielversprechend

Während substantivische Kookkurrenzen aus der Liste nur Indikatoren für die Lesart sein können und eine ungünstige Ambiguität mit der Lesart 'modern, aktuell' verkörpern, dienen die umgebenden Kontexte dieser Indikatoren der Ermittlung von paradigmatischen Konstruktionen, deren unterschiedlich ausgeprägte Muster die beiden Lesarten voneinander trennen.

³² Als weiteres Antonym wird *it sucks* interpretiert, ein amerikanischer Ausdruck, der persönliche Missbilligung oder Abwertung zum Ausdruck bringt, aber noch nicht in der deutschen Sprache lexikalisiert ist.

3.2.3 Eigenschaftsprädikatore

Eigenschaftsprädikatore treffen Aussagen über Eigenschaften und Qualitäten von Lebewesen, Sachen oder Abstraktbegriffen. Eigenschaften sind zugeschriebene Merkmale und gelten als grundsätzlich unveränderlich. Im Kontrast zu Zuständen sind sie keinem Zeitverlauf unterworfen.

3.2.3.1 'schlicht, einfach, nüchtern'

Eine aufgrund weniger Belege und Indikatoren nur schwer identifizierbare Lesart stellt die Lesart 'schlicht, einfach, nüchtern' dar. Es verwundert daher auch nicht, dass diese Lesart nicht im Duden GWDS (2000) eingetragen ist. Mit ihr wird auf Kontexte Bezug genommen, in denen der Mangel an Verzierung und Ausschmückung von Gegenständen oder auch von Ereignissen beschrieben wird. Auch wenn das Kookkurrenzverzeichnis selbst keinen Indikator, der auf diese Lesart hinweist, enthält, ergaben sich bei genauer Analyse Beweise für das Ansetzen dieser Lesart. Die Beweise wurden durch die Verifizierungsanalyse der Kollokatoren gefunden, die als Indiz für vermeintlich andere Lesarten interpretiert wurden. Von daher muss man hier von einem unerwarteten Ergebnis sprechen. Diese zufällige Entdeckung deckt die Grenzen des computergestützten Verfahrens auf und unterstreicht die Notwendigkeit einer systematischen manuellen Untersuchung aller Kollokatoren und deren Kontexte, um Unerwartbares entschlüsseln zu können.

Im Zusammenhang mit dem Kollokator *heiß* wurde der folgende Beleg entdeckt, der auf die Lesart 'schlicht, einfach, nüchtern' aufmerksam machte:

Beispiel:

Es gibt neben ein paar Alibi-Reportagen genug Bilder, Kästchen, „heisse Sprüche und **coole Bekenntnisse**“ zu so existentiellen Problemen wie Geburt, Zeugung oder Mitropa-Zugrestaurants. (MAI.12776 Züricher Tagesanzeiger, 29.05.1997, S. 65)

Im Zusammenhang mit dem Mitspieler *Jazz* fand sich folgendes Textbeispiel:

Entspannt groovt das Septett auf ihrem Debüt über sechs Stücke, garniert sich mit Jazzrock- und Funk-Glasur und überzeugt durch eine angenehm undeutsche Art Jazz zu spielen. Keine verquastenen Soli ohne Herztöne, kein Unisono-Body-Building, vielmehr **coole Schlichtheit** gibt ihrem Debüt Profil. Gelungenes Zusammenspiel auf fortlaufenden Beats bestimmt die Architektur. Die unaufgeregte Art führt zwar auch Momente lang zu Dämmer Schlaf, prägt aber einen eigenen Stil. (T93/NOV.49270 die tageszeitung, 05.11.1993, S. 24)³³

Aufgrund der wenigen Hinweise auf diese Lesart lassen sich kaum Aussagen zu typischen paradigmatischen und syntagmatischen Strukturen treffen. Hier ist die Introspektion unerlässlich. So können mittels linguistischer Kompetenz Prüfanfragen gestellt werden, die einen Kontext bieten, in dem erwartbare Mitspieler auftauchen könnten.

³³ Vermutlich ist diese Lesart durch Konzeptverschiebung der Zustandsbeschreibung 'kalt, kühl' hervorgegangen.

Der Wirt des beliebten Altsttter Lokals «alte Schmiede» gestaltete seinen zweiten Betrieb in der Art des «Conrad Sohm» in Dornbirn oder der Feldkircher «Sonderbar». Er machte mit der «U-Bar» also nicht auf «schick», sondern auf «einfach, schlicht, cool». (A00/JAN.03876 St. Galler Tagblatt, 12.01.2000)

Auf diese Weise werden Lexeme wie *einfach, schlicht, nchtern* als koreferenzielle Mitspieler mit hnlicher Referenz und *schick, schwungvoll, geschnt, kitschig* als antonymische Partner ermittelt.

Im Laufe der Jahrzehnte entwickelte sich der Landweg zwischen Berlin und Potsdam zum herausragenden Straenzug. Smtliche Baustile, die vom Brgertum in den letzten 100 Jahren kreiert worden sind, versammeln sich dicht gedrngt an der ehemaligen Chaussee: Eklektizismus, Klassizismus, Grnderzeit, Jugendstil und nchterner Zweckbau des zwanzigsten Jahrhunderts. Schwungvoll, schn oder geschnt, kitschig und cool. (B99/907.62728 Berliner Zeitung, 10.07.1999; Flaniermeile mit Wohnkultur, S. 49)

In diesem Fall kann die Korpusrecherche nur als Belegsuche und Prfinstrumentarium fr linguistische Erwartbarkeit dienen, nicht aber zur primren Identifikation der Lesart.

3.2.3.2 'modern'

Eine Prdikation, in der eine weitere Eigenschaft beschrieben wird, ist in der Lesart 'modern' impliziert. Bezug wird auf modernes Zeitgefhl und auf im Trend befindliche Objekte/Wesen genommen. Es wurde bereits hervorgehoben, dass die Abgrenzung zur Lesart 'super, sehr gut gefallend, interessant' als Einstellungsprdikator nicht gewhrleistet werden kann. Was modern ist, gilt oft als wohlgefallend und interessant, und hufig erfolgt eine Aufwertung. Aber dieser Rckschluss ist nicht immer zulssig, und nicht immer kann von einer positiven Bewertung in Zusammenhang mit Modernitt die Rede sein. Einige der Kookkurrenzindikatoren tauchen daher bei beiden Lesarten auf, der Interpretationsspielraum ist gro, eine genaue Lesartenzugehrigkeit kann nicht identifiziert werden.

BelegNr	Gamma	Kollokatoren	Hufigkeit
2676+46:	368	Sonnenbrille	46
5640+13:	106	Outfit	13
5770+6:	90	Klamotte	6
6036+8:	67	Design	8
6044+24:	67	Musik	24

Durch die Verifizierungsanalyse lassen sich die Kollokatoren nach thematischen Kontexten wie Musik und Mode gruppieren. Das geht bereits aus der Kollokationsliste hervor, aber auch bei der Tiefenanalyse erscheinen immer wieder Mitspieler, die diesen Bereichen zuzuordnen sind. So wurden aus dem Bereich der Mode folgende Mitspieler ermittelt: *Anzug, Brille, Design, Frisur, gestylt, Jeans, Jacke, Klamotten, Marken, Outfit*. Aus dem Bereich der Musik ergaben sich folgende Wrter: *Album, Beat, Groove, HipHop, Musik, Rap, Song, Sound, Vibes*. Sie bilden thematische Felder, die beiden Lesarten zugeordnet werden knnen. Die thematische Gruppierung von substantivischen Mitspielern nach Referenzbereichen kann hier nicht als hinreichendes Abgrenzungskriterium fungieren. Daher mssen zustzlich relationale Strukturen innerhalb der Paradigma-

tik untersucht werden. Das folgende Beispiel aus der Kolloaktionsliste illustriert einen koreferenzielle Mitspieler:

BelegNr	Gamma	Kollokatoren	Häufigkeit
5791+1:	88	trendy	1

Coole Klamotten, **trendig**e Sonnenbrillen (die auch während des Spiels nicht abgenommen werden), austrainierte und gebräunte Körper gehörten genauso zu der Szenerie wie knallharte Schmetterbälle, tolle Hechtsprünge und rasante Ballwechsel. (R98/JUN.43786 Frankfurter Rundschau, 03.06.1998, S. 38)

Nach der Analyse der KWIC-Ebene und der Volltextebene lassen sich eine Reihe relationaler Strukturen erkennen und in einem koreferenziellen Feld bündeln:

aufgeklärt, chic, clever, dynamisch, frisch, funky, großzügig, krass angesagt, heavy, hochmodern, heiß, hip, hot, jung, kess, mega in, modern, new, schick, streetwise, trendy, trendig, urban

Die Merkmalskomponenten 'Zeit' und 'Aktualität', die in vielen der Lexeme als inhärente Merkmale enthalten sind, kommen auch bei den Antonymen zum Ausdruck: *ein edler Klassiker, out, uncool, unzeitgemäß*. Ein Vergleich der paradigmatischen Muster zwischen Lesart 'modern' und 'super, sehr gut gefallen, interessant' zeigt, dass sich die Netze nicht decken und daher nur die Paradigmatik als hinreichendes Disambiguierungskriterium zu Hilfe gezogen werden kann. Der Entschluss, eine weitere Lesart zu definieren, ist daher gerechtfertigt, wenn auch die Interpretation einiger Kollokatoren und deren Kontexte umstritten und eine klare semantische Einordnung aller Belege in die einzelnen Lesarten nicht immer möglich ist.

3.2.4 Referenzwörter

Als Referenzwort werden Wörter klassifiziert, die auf einen einzigen Gegenstand der Welt Bezug nehmen, also prototypischerweise Eigennamen. Diese Lexeme haben referierende und kaum prädikative Funktion. Referenzwörter sind in hoher Zahl in der Kookkurrenzliste abgebildet. Der folgende Ausschnitt erfasst die Gesamtzahl längst nicht befriedigend:

BelegNr	Gamma	Kollokatoren	Häufigkeit
121+116:	2151	J.F	116
343+6:	1634	LL Epp	6
535+20:	1569	Jazz Bebop	20
555+9:		Jazz Jazzbistro Gambrinus	9
564+176:		Jazz	176
740+169:	1381	Typ	169
2034+6:	878	Ecstasy Ballett	6
2040+12:		Ecstasy ausverkaufen	12
2052+52:		Ecstasy	52
2198+15:	755	daddy basieren Baby Film	15
2213+3:		daddy Breeze Jazzbistro sweet Ga	3
2223+1:		daddy Breeze sweet	1
2224+27:		daddy	27
2251+11:	658	Kid wait can	11
2262+79:		Kid	79
3243+2:	306	Mister Planke 16.45	2
3256+20:		Mister	20
4899+1:	259	Bebop Swing Hardbop	1
4900+1:		Bebop Swing	1
4901+5:		Bebop	5

Cool als Teil von Eigennamen deutet auf eine Konzeptübertragung einzelner Lesarten auf Personen oder Dinge, denen bestimmte im Lexem *cool* implizierte Zustände, Eigenschaften oder deren Bewertungen zugeschrieben werden, hin. Diese Angaben sind nicht in den Dudenwortartikel eingegangen. Bei den Referenzwörtern handelt sich um unterschiedliche Arten von Eigennamen, z. B. Personennamen, Titel für Musik- oder Ballettstücke, Namen für Musikrichtungen, Namen von Institutionen und Produkten. Eine detaillierte Volltextsuche ergibt hier keinen großen Zugewinn an semantischer Information zur Lesartendisambiguierung, da sich Eigennamen semantisch anders als andere satzsemantische Klassen verhalten. Die Referenzwörter werden als eigenständige Klasse und damit als gesonderte Lesart behandelt. In der folgenden Tabelle sind exemplarisch die unterschiedlichen Referenzwörter mit Belegen dargestellt:

Referenzwort	Beispiel	Quelle
Personenname	George Bush und Jimmy Carter, Football Star John Elway und Rap-Sänger LL Cool J.	B97/711.08337 Berliner Zeitung, 11.11.1997, S. 14
Name für Jazz-richtung	Vielleicht spielen die Norweger auch deswegen im "Moments", home of Urban Style, Cool Jazz und Latino House	T99/FEB.07930 die tageszeitung, 19.02.1999, S. 23,
Ballettname	Einführungsmatinée zur Ballett-Premiere "Cool Ecstasy" (freier Verkauf).	M98/801.02093 Mannheimer Morgen, 10.01.1998
Agenturname	Die Agentur "Cool Kids" eröffnet am Montag an der Anklamer Straße eine Schauspielschule ...	B00/004.27402 Berliner Zeitung, 03.04.2000, S. 24
Name für Rasierapparat	Philishave hat den Cool Skin im Frühjahr lanciert und bietet jetzt bereits ein De-Luxe-Modell an [...].	E98/OKT.26915 Züricher Tagesanzeiger, 20.10.1998, S. 75

4. Schlussbemerkungen

In diesem Artikel wurde der Versuch unternommen zu zeigen, inwieweit ein linguistisches Disambiguierungsmodell, Korpora und ein Analysetool für die lexikografische Lesartenidentifikation geeignet sind. Die Resultate belegen, dass *ellexiko* durch *Cosmas* und Korpora, die die Arbeitsgrundlage der Datenextrahierung bilden, sowohl zu umfangreicheren semantischen Beschreibungen als auch zu zusätzlichen Lesarten gegenüber dem Duden GWDS (2000) kommt. Die Analyse hebt hervor, dass trotz einiger Schwierigkeiten das Instrumentarium Kollokationsanalyse und die damit verbundenen Ebenen der Verifizierungsanalyse aussagekräftige Ergebnisse liefern. Sinclair betont zu Recht, dass durch umfangreiche Datenmengen und objektive Strukturierungen eine größere Präzision bei der Analyse von sprachlichen Daten erreicht werden kann:

The general conclusion is that much greater precision can be gained in language work by studying instances which are gathered objectively. (Sinclair 1991, S. 51).

Der Mehrgewinn gegenüber dem Duden GWDS (2000) besteht aber nicht nur im Erkennen zusätzlicher Lesarten, sondern auch in der methodischen Absicherung der bereits vorhandenen Lesarten. *Ellexiko* hat ein lexikografisches Verfahren der Disambiguierung entwickelt, das theoretische und empirische Aspekte vereint. Mithilfe eines mehrstufigen linguistischen Klassifikationsmodells werden die unterschiedlichen propositionalen Verwendungsweisen eines Lexems linguistisch gruppiert. Die Identifikation von Lesarten ist daher nicht nur ein praktischer, korpusbasierter und softwaregestützter, sondern darüber hinaus auch ein modellgeleiteter Prozess.

Von besonderem Nutzen für die lexikografische Arbeit sind bei der Lexemanalyse die Kollokations-, die KWIC- und die Volltextebene. Auf jeder dieser Ebenen können semantische Informationen extrahiert werden, diese werden aber unterschiedlich auffällig abgebildet. Die Ergebnisse werden umso detaillierter für das semantische Gesamtbild eines Lexems, je tiefer in die Ebenen eingedrungen wird. Die Kollokationsebene bietet durch ihre Hierarchisierung und Systematisierung von kontextuellen Mitspielern den ersten Zugang zur Semantik, insbesondere die ersten Indikatoren für unterschiedliche satzsemantische Klassen, Referenz und Lesarten. Die Betrachtung der Indikatoren in der Tiefenanalyse führt zu Einblicken in paradigmatische und syntagmatische Kookkurrenzen, welche sowohl inhaltlich-thematische als auch strukturelle Informationen zur Bedeutung und Verwendung des Stichworts liefern. Sämtliche Kookkurrenzen formen schließlich paradigmatische Muster, determinieren syntagmatische Kohäsion und bieten die für die Lexembeschreibung notwendigen pragmatischen Beschreibungen. Des Weiteren sind Lexikografinnen durch das korpusgestützte Verfahren in der Lage, objektivere Aussagen über Sachgebiete oder Domänen, in denen bestimmte Wörter auftauchen, zu treffen. Angaben zu Mitspielern, die thematisch miteinander verwandt sind oder bestimmte semantische Funktionen im Verhältnis des Satzes ausüben, wie z. B. semantische Rollen, können ermittelt werden. Ganze Felder von Mitspielern lassen sich systematisch extrahieren und gruppieren. Mittels der Resultate, die das Analyse- und Rechercheinstrument liefert, werden Einblicke in die linguistische Erwartbarkeit an Mitspielern, also in assoziative Kookkurrenzen, gewonnen. Die Verifizierungsanalyse informiert zusätzlich über Verwendungssituationen, die außerhalb der Erwartbarkeit (untypischer

Mitspieler) liegen, präsentiert also Ergebnisse, die durch bloße Introspektion nicht erfasst worden wären. Das große Verdienst von *Cosmas* liegt in der Erfassung von Mas-sendaten, deren Vorstrukturierung durch statistische Auswertung von Frequenz und Sig-nifikanz und in seinem empirischen Zugang zur Sprache, d.h. ohne Erwartung an Kon-texte heranzutreten. Scholze-Stubenrecht (2001) verweist bereits auf die Vorteile der Kollokationsanalyse, auch wenn sie in der Dudenredaktion noch nicht zum Einsatz kommt:

Die Erstellung von Kollokationsindizes aus Volltextkorpora wird hier einige Fortschritte bringen, aber diese Entwicklung steht noch in ihren Anfängen, und es ist abzusehen, ob sie in absehbarer Zeit zu der Qualität der von Menschen vorstrukturierten und vorselektierten Belegsammlung führen kann. (Scholze-Stubenrecht 2001, S. 62)

Cosmas ist ein Beispiel dafür, dass Analysetools die lexikografische Arbeit vorteilhaft verändern können.

Eine vollständige Ermittlung von Lesarten und semantischer Beschreibung von Inhalten und Strukturen kann durch *Cosmas* nicht in jedem Fall gewährleistet werden. Die linguistische Kompetenz kann nicht durch Statistik ersetzt werden, da der Interpretationsfrei-raum bei einigen Lexemen groß sein kann. Wie wichtig linguistische Kompetenz ist, zeigt auch die Tatsache, dass man erkennen muss, inwieweit Kollokatoren *nicht* für eine Lesartenidentifikation in Frage kommen. Irrelevante Partner³⁴ müssen erkannt und inter-pretiert werden. Die Entscheidung, wie mit welcher Korpusinformation umzugehen ist und welche Schlussfolgerungen zu ziehen sind, bleibt dem Menschen vorbehalten. Das statistische Analysewerkzeug ist nicht in der Lage, Kookkurrenzen nach Lesarten zu gruppieren. Die Suchoptionen und optimalen Parametereinstellungen, die dazu dienen, bestimmte linguistische Aspekte ein- oder auszuschließen, müssen vorher eruiert wer-den, um gezielt Information aus den Korpora zu extrahieren. Nur durch wohldefinierte Suchanfragen und ein geübtes Auge sind Lexikografinnen in der Lage, dieses Instru-mentarium gewinnbringend zu nutzen. Man muss daher ein Gespür für das Sortieren entwickeln, die Vorstrukturierung der Indikatoren richtig interpretieren und sich dann selbst ein systematisches Vorgehen zur Erkenntnisgewinnung erarbeiten. Es bleibt im-mer die Aufgabe der Lexikografinnen, diese Korpusdaten zu analysieren und zu inter-pretieren. Aber es ist erkennbar, dass modellgeleitete und computergestützte Disambigui-erung die lexikografische Qualität der Lesartenidentifikation verbessern kann. Compu-tergestützte Verfahren liefern auch neue empirische Erkenntnisse für andere Gebiete und sollten daher für LinguistInnen/Lexikografinnen verschiedenster Bereiche in Zukunft von Interesse sein und eine noch größere Rolle spielen.

5. Literatur

Belica, Cyril (1995-2002): Statistische Kollokationsanalyse und Clustering. COSMAS-Korpusanalyse-modul. Institut für Deutsche Sprache, Mannheim.

Haß-Zumkehr, Ulrike (2002): Das Wort in der Korpuslinguistik – Chancen und Probleme empirischer Lexikologie. In: Ágel, Vilmos (Hrsg.): Das Wort: seine strukturelle und kulturelle Dimension. Festschrift für Oskar Reichmann zum 65. Geburtstag. Tübingen: Niemeyer, S. 45-70.

³⁴ Z. B. Funktionswörter, die keinen semantischen Beitrag zur Disambiguierung leisten.

- Haß-Zumkehr, Ulrike (erscheint voraussichtlich 2003): Das Projekt Wissen über Wörter des Instituts für Deutsche Sprache. In: Scharnhorst, Jürgen (Hrsg.): Sprachkultur und Lexikographie. Frankfurt: Peter Lang. (Aus dem Manuskript).
- Jones, Steven (2002): *Antonymy – A corpus-based perspective*. London: Routledge.
- Lawson, Ann (2000): Die schöne Geschichte: a corpus-based analysis of Thomas Mann's *Joseph und seine Brüder*. In: Dodd, Bill (Hrsg.): *Working with German Corpora*. Birmingham: The University of Birmingham Press, S.161-180.
- Polenz, Peter v. (1988): *Deutsche Satzsemantik. Grundbegriffe des Zwischen-den-Zeilen-Lesens*, (= Sammlung Götschen 2226), Berlin/New York: de Gruyter.
- Pustejovsky, James (1995): *The Generative Lexicon*. Cambridge, Mass.: MIT Press.
- Reichmann, Oskar (1989): *Frühneuhochdeutsches Wörterbuch. Bd.1 Einführung, a – äpfelkern* Hrsg. von Anderson, Robert R.; Goebel, Ulrich; Reichmann, Oskar. Berlin/New York: de Gruyter.
- Schölze-Stubenrecht, Werner (2001): Das Internet und die korpusgestützte praktische Lexikographie. In: Korhonen, Jarmo (Hrsg.): *Von der mono- zur bilingualen Lexikographie für das Deutsche*. (Finnische Beiträge zur Germanistik Bd. 6). Frankfurt: Peter Lang, S. 43-64.
- Schölze-Stubenrecht, Werner (2002): Die Auswahl der Einträge ist äußerst beliebig. In: *Zeitschrift für Sprachwissenschaft*. Heft 27/2, S. 225-248.
- Sinclair, John (1987): Introduction. In: Sinclair, John (Hrsg.): *Collins Cobuild English Dictionary*. London/Glasgow: Collins. S. XV-XXIII.
- Sinclair, John (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Steyer, Kathrin (2000): Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikografische Möglichkeiten. In: *Deutsche Sprache* 2/00, S. 101-125.
- Steyer, Kathrin (2002): Wenn der Schwanz mit dem Hund wedelt. Zum linguistischen Erklärungspotenzial der korpusbasierten Kookkurrenzanalyse. In: Haß-Zumkehr, Ulrike/Kallmeyer, Werner/Zifonun, Gisela (Hrsg.): *Ansichten zur deutschen Sprache. Festschrift für Gerhard Stickel zum 65. Geburtstag*. (= Studien zur deutschen Sprache 25). Tübingen: Niemeyer, S. 215-236.
- Strauß, Gerhard (1989): Angabe traditioneller Wortarten oder Beschreibung nach funktionalen Wortklassen im allgemeinsprachigen Wörterbuch? In: Hausmann, Franz Josef; Reichmann, Oskar; Wiegand, Herbert Ernst; Zgusta, Ladislav (Hrsg.): *Wörterbücher: ein internationales Handbuch zur Lexikographie*. (Handbücher zur Sprach- und Kommunikationswissenschaft Bd. 5.1). Berlin/New York: de Gruyter, S. 788-796.
- Werner, Reinhold (1989): Probleme der Anordnung der Definitionen im allgemeinen einsprachigen Wörterbuch. In: Hausmann, Franz Josef; Reichmann, Oskar; Wiegand, Herbert Ernst; Zgusta, Ladislav (Hrsg.): *Wörterbücher: ein internationales Handbuch zur Lexikographie*. (Handbücher zur Sprach- und Kommunikationswissenschaft Bd. 5.1). Berlin/New York: de Gruyter, S. 917-930.

Wörterbuch

- Duden GWDS (2000): *Das große Wörterbuch der deutschen Sprache*. 10 Bände auf CD-Rom. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.

Dr. Petra Storjohann
 Institut für Deutsche Sprache
 Postfach 101621
 D-68016 Mannheim
 e-mail: storjohann@ids-mannheim.de